

**COMMUNICATIONS SYSTEM AND METHOD UTILIZING  
CENTRALIZED SIGNAL PROCESSING**

INVENTORS

**JAMES H. PARRY**  
P.O. Box 557  
Saratoga, CA 95071

**PETER HSIANG**  
P.O. Box 485  
Cupertino, CA 95015-0485

002160" 50209950

# COMMUNICATIONS SYSTEM AND METHOD UTILIZING CENTRALIZED SIGNAL PROCESSING

INVENTORS

**JAMES H. PARRY**  
**PETER HSIANG**

5

## BACKGROUND

### FIELD OF THE INVENTION

This invention pertains in general to telephony and in particular to performing centralized acoustic signal processing on audio signals received from terminals engaged in a communications session.

10

### BACKGROUND ART

Hands-free two-way audio communications systems, such as speakerphones, utilize both a microphone and a speaker. The microphone transmits speech and other sounds from the local terminal to remote terminals while the speaker emits sounds received from remote terminals. In a typical speakerphone system, the speaker and microphone are located in close proximity and sounds produced by the speaker are picked up by the microphone. Without signal processing, therefore, a feedback loop is easily created between the speaker and microphone. This feedback can cause the speaker to emit an undesirable "howling" noise and cause the remote terminals to hear echoes.

15

20

One simple technique for eliminating feedback is to provide half-duplex switching where only the microphone or the speaker is active at any given instant. In a typical half-duplex system, the speaker is active until a sound is detected at the microphone. Then, the speaker becomes inactive and the microphone becomes active for the duration of the

sound. Half-duplex systems have many inherent problems, not the least of which is that a slight noise may unintentionally cause the speaker to cut out. As a result, it is often difficult to conduct a normal conversation using a speakerphone with a half-duplex switching system.

5           More sophisticated audio communications systems use complicated adaptive techniques to reduce background noises as well as to regulate gain in the audio channel and eliminate feedback. These techniques identify selected acoustical situations, such as “doubletalk” or “voice not present,” and use these identifications to control the rate of adaptation of the signal conditioning methods. Adaptive acoustic echo cancellation  
10 (AEC), for example, is performed at the speakerphone and utilizes a sample-by-sample copy of the signal going to the speaker as the basis for an estimate of the echo returning through the microphone. The estimated echo is subtracted on a sample-by-sample basis in an attempt to separate out only that portion of the microphone signal due to sounds coming from sources other than the speaker.

15           Other signal processing techniques may also be performed at the terminal to improve the quality of the audio signal. For example, frequency shifting is sometimes used to further attenuate loop gain at a particular frequency and thus avoid howling. In addition, a noise reduction algorithm can be used to estimate a frequency dependent profile of the noise floor and attenuate sounds which are temporarily near or below that  
20 noise floor. A voice-gated automatic gain control (AGC) can also be used to isolate times during which local speech is thought to be present and then adjust the signal gain so that the speech is near a predetermined level when considered on average.

These solutions can work reasonably well, but the software and hardware for implementing these solutions is integrated into the speakerphone. Thus, the software and hardware must be replicated for each speakerphone and the total cost of the solution depends upon the number of speakerphones in existence. In addition, each speakerphone must be designed and built to some pre-selected level of quality and can not be easily tuned to a particular use or environment. Similarly, it is not easy to change or upgrade the solution implemented by the speakerphone.

Therefore, there is a need in the art for a solution that provides effective signal processing to an audio communications system but does not have costs that scale with the number of speakerphones. Preferably, this solution would also allow the signal processing performed for each speakerphone to be easily tuned or upgraded for a particular use or environment.

#### DISCLOSURE OF THE INVENTION

The above needs are met by an audio communications system and method that performs centralized digital signal processing. Since there is no need to provide digital signal processors in the terminals, the cost of the terminals is reduced. Moreover, the present invention allows easy tailoring and upgrading of the capabilities of the communications system.

In an embodiment of the present invention, a plurality of terminals are coupled to a processing switch. The terminals can be, for example, dedicated speakerphones, desktop handsets, or personal computers with audio capabilities. The terminals can be coupled to the switch via wired and/or wireless links.

The processing switch provides support for audio communications. The centralized signal processing capabilities of the switch are provided by a communications support module (CSM). Under direction of the CSM, the switch can receive, process, and send data to/from multiple terminals simultaneously. In addition, the switch can support multiple simultaneous communications sessions, where each session contains two or more terminals engaged in communications.

In one embodiment, the CSM develops and stores room models having data sets representing the acoustic properties of the terminals and/or the terminals' environments. A room model can also hold state information reflecting signals previously sent to the associated terminal.

The CSM is also supported by one or more signal processor (SP) modules. The SP modules provide a pool of processing resources from which the CSM can draw to process audio signals received from or sent to the terminals. In one embodiment, the CSM can dynamically allocate and deallocate SP resources in response to the overall system load on the processing switch or the characteristics of the particular audio signals received from the terminals. Thus, if relatively few communications sessions are being conducted on the switch, more processing power can be allocated to particular signals than if many sessions are occurring simultaneously.

In operation, the processing switch receives audio signals from one or more of the communications terminals. If necessary, the CSM decodes the received audio signals into a format suitable for further processing. Then, the CSM processes the decoded signals according to desired acoustical procedures.

The CSM uses the SP modules to perform signal processing on the received audio signals. The types of signal processing available include acoustic echo cancellation (AEC), automatic gain control, noise reduction, and signal mixing. Preferably, the CSM uses the room models to determine the optimal signal processing to perform on the audio signals. In one embodiment, the CSM can dynamically allocate SP power to particular audio signals in response to difficult room environments such as rooms having long reverberation times or high noise levels.

The CSM is also preferably supported by a mixing module. The mixing module formulates an output mix for each of the terminals. In general, each terminal's output mix is comprised of a mix of the received audio signals from the other terminals. If only one terminal is producing an audio signal, then the formulated output mix merely contains the processed audio signal from the SP module and the terminal producing the audio signal does not receive the output mix.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a block diagram of a prior art audio communications system;

FIGURE 2 is a high-level block diagram of an audio communications system according to an embodiment of the present invention; and

FIGURE 3 is a flowchart illustrating the operation of the communications support module according to an embodiment of the present invention.

**DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS**

FIG. 1 is a block diagram of a prior art switched audio communications system 100. A plurality of terminals 110A-C are coupled to a central switch 112 via communications links 114A-C. Each terminal 110, of which terminal 110A is representative, includes a microphone 116A and a speaker 118A. As used herein, a speaker is any device that produces audible motion in response to an electrical signal and a microphone is any device that produces an electrical signal in response to audible motion. The communications links 114 carry sounds picked up by the microphone 116 and to be played by the speaker 118 to/from the switch 112.

The terminal 110 contains a signal processor (SP) 120 with which the terminal performs acoustic echo cancellation (AEC). The AEC removes the speaker 118 sounds that are picked up by the microphone 116. The switch 112 performs switching and routing of audio signals by determining from which terminal an audio signal is being received and to which terminal(s) the audio signal should be sent.

FIG. 2 is a high-level block diagram of an audio communications system 200 according to an embodiment of the present invention. A plurality of terminals 210A-D are coupled to a processing switch 212 via communications links 214A-D. The terminal types can be heterogeneous or homogeneous. In one embodiment, the terminals include: dedicated speakerphones, desktop handsets with or without speakerphone capabilities, cellular phones, desktop personal computer (PC) systems with audio capabilities, video conferencing systems with audio capabilities, etc. Each terminal 210, of which terminal 210A is representative, preferably includes a microphone 216A and a speaker 218A. Unlike in the system of FIG. 1, the terminals 210 need not, and preferably do not, contain

a SP for performing AEC or other complex signal processing. However, the terminals 210 may have SPs for performing other functions, such as analog-to-digital and digital-to-analog-signal conversions and, optionally, data encoding/decoding.

The communications links 214 carry audio data representative of sounds picked up by the microphone 216 and to be played by the speaker 218 to/from the processing switch 212. The communications links 214 may be wired or wireless. Moreover, the links 214 may include dedicated private links, shared links utilizing a publicly-accessible telephone network, and/or links using a public or private data communications network such as the Internet. Data traveling over the links 214 may pass through one or more switches or link types before reaching the processing switch 212 or terminal 210, although a preferred embodiment of the present invention treats a link passing through multiple links and switches as a single logical link. The data carried by the communications links 214 can be digital and/or analog. If the data is digital, it is preferably transmitted as a series of discrete data packets, such as Internet protocol (IP) packets. In one embodiment, the digital data is encoded into a compressed format.

The processing switch 212 switches and routes communications among the terminals 210 and provides signal processing as described herein. The switch 212 can be, for example, a private branch exchange (PBX) located at a business or other entity, a publicly-accessible switch operated by a telephone company or other entity providing audio communications, or an Internet server supporting Internet telephony. Thus, the term "processing switch" includes any device capable of providing the switching and processing functionality attributed to the processing switch 212 described herein.



In one embodiment, the processing switch 212 comprises a dedicated computer system having one or more central processing units (CPUs), a random access memory (RAM), a read-only memory (ROM), a storage device such as a hard drive, switching hardware and software, and other hardware and software components for providing switch functionality as is known in the art.

Aggregations of machine-executable code, data, circuitry, and/or data storage areas in the processing switch 212 for performing a specific purpose or purposes are referred to as "modules." Different modules may share common code, data, and/or circuitry. The modules include, for example, modules for receiving and sending data, a digital to analog converter (DAC) module, and an analog to digital converter (ADC) module. The switch 212 can perform multiple tasks simultaneously by allocating a subset of available modules, processors, processing time, or other resources to particular tasks.

According to an embodiment of the present invention, the processing switch 212 has a communications support module (CSM) 220 for supporting communications utilizing shared and centralized signal processing. Under direction of the CSM 220, the switch 212 receives the data provided by the communications links 214, processes the data using centralized resources and modules, and provides the data to the appropriate terminals.

In one embodiment, the CSM 220 associates a room model 222 with each terminal engaged in a communications session. The room model 222 holds room model information including data and/or filters representing the acoustic properties of the terminal 210 and/or the environment surrounding the terminal. For example, in one embodiment the room model 222 holds data representing the reverberation characteristics

of the room in which the terminal 210 is located. The room model 222 can also hold data representing an amount of background noise present at the terminal, an amount of automatic gain control (AGC) to be applied to audio data received from the terminal, types of noise reduction to be applied to signals received from or sent to the terminal, or any other information useful for supporting signal processing to be performed on data received from or sent to the associated terminal 210. Room model information held in the room model 222 can also include state information indicating signals previously sent to the associated terminal 210. In one embodiment of the present invention, the resources utilized by the room model 222, such as memory and processing power, are dynamically allocated. For example, if a room is found to have a long reverberation time, a larger memory is allocated to the room model 222 in order to store the sequencing information.

In one embodiment of the present invention, the room model 222 is adaptively developed while the terminal 210 is used in a communications session. This technique is preferred because the room model 222 may change during the course of the communications session. For example, a participant may switch the terminal 210 from a handset to a speakerphone midway through a conversation. Alternatively, the room model can be generated by sending a series of test signals to the terminal at different points during the session.

In one embodiment, room models 222 are persistently stored in the processing switch 212 and retrieved from storage each time an associated terminal 210 is used in a communications session. For example, if the set of terminals interacting with the switch 212 is finite, known, and has relatively constant acoustic properties, it may be more efficient to store a persistent room model for each terminal 210. In contrast, if the switch

212 is a server on the Internet and the set of terminals is unknown and practically infinite, it may be more efficient to generate a new room model each time a terminal establishes a connection with the switch.

The CSM 220 is supported by one or more SP modules 224. The SP modules 224 preferably operate in the digital domain, but may also provide analog processing. Taken together, the SP modules 224 provide a pool of processing resources from which the CSM 220 can draw. During times when the system load is light (i.e., relatively few terminals are engaged in audio communications), correspondingly greater processing resources are available for use with each supported terminal 210. Likewise, when the system load is heavy (i.e., a relatively large number of terminals 210 are engaged in audio communications), the processing resources available for each terminal are diminished.

In a preferred embodiment of the present invention, the CSM 220 monitors the received audio signals and the processing needs of the switch 212 and dynamically allocates and deallocates the processing resources of the SP modules 224 to the signals in order to provide maximum benefit to the active terminals (i.e., terminals having microphones generating audio signals). In one embodiment, this monitoring is performed by analyzing the digital packets received from the terminal 210 over the communications links 214. If the data in the packets forming the signal are determined to represent audio data generated by a person speaking, SP resources allocated to that signal. The amount of resources allocated depends on the amount of SP resources available in the pool. If, on the other hand, the CSM 220 determines from the packets that no one is speaking at the terminal, the CSM preferably digitally mutes the signal and allocates relatively few SP resources to it.

The CSM 220 is preferably supported by a mixing module 226. The mixing module 226 formulates an output mix for each of the terminals 210. In general, each terminal's output mix is comprised of a mix of the received audio signals from the other terminals. If only one terminal is producing an audio signal, then the formulated output  
5 mix merely contains the processed audio signal from the SP module 224 and no output mix is sent to the terminal producing the audio signal. In one embodiment, the mixing module 226 is merely a logical construct formed when the CSM 220 uses the SP modules 224 to perform mixing.

FIG. 3 is a flowchart illustrating the operation of the CSM 220 when providing  
10 real-time centralized signal processing according to one embodiment of the present invention. For purposes of example, assume there are N terminals engaged in a communications session (e.g., N speakerphones engaged in a conference call). N may be a subset of the total number of terminals coupled to the processing switch 212 and there may be multiple communications sessions ongoing simultaneously. As is understood in  
15 the art, there are many different ways for the initial communications session to be established. For example, all of the participants can call into the switch 212 or one participant can initiate the session by calling the other participants.

At step 310, the processing switch 212 receives audio signals from one of the terminals 210 via the communications links 214. At any given time, one or more of the  
20 terminals 210 can produce audio signals and send the signals to the switch 212. Thus, the switch 212 can simultaneously receive audio signals from multiple terminals. Audio signals received by the switch 212 are made available to the CSM 220 and SP modules 224 in the switch.

Next, the CSM 220 uses the SP modules 224 to decode 312 the signals received via the communications links 214, if necessary. As mentioned above, the audio signals can be digitally encoded to compress the signal, detect errors, and/or provide other benefits. The CSM 220 decodes and/or decompresses the audio signals and stores the  
5 signals in a format suitable for performing further processing. Since different terminals 210 may use different encoding/decoding methods, the CSM 220 preferably simultaneously supports multiple encoding/decoding methods.

The CSM 220 uses the SP modules 224 to process 314 the decoded signals according to desired acoustic procedures. The types of processing 314 that can be  
10 performed on the decoded audio signals include: AEC, AGC, noise reduction, and mixing. Of course, this list is not exclusive and any type of desired processing can be performed on the signals. This processing may utilize the room models 222 and, in addition, may update the room models 222.

For adaptive AEC, the CSM 220 determines which terminals are active. Then, the  
15 CSM 220 uses the associated room model 222 to process the signal from that terminal and remove the echo caused by the microphone 216 picking up sounds from the speaker 218. In one embodiment, digital sample values of the audio signals previously sent to the terminal 210 are stored in the associated room model 222. The stored digital sample values are used to estimate the echo returning through the microphone 216 of that  
20 terminal and the estimated echo is subtracted on a sample-by-sample basis from the received audio signal. In one embodiment, the outgoing packets bear sequencing information, such as time sequence tags, which is used to determine a correlated time sequence on the returned packets of audio information. In other words, the time sequence

tags are used to maintain alignment between the samples delivered to the terminal's speaker 218 and the samples received from the terminal's microphone 216.

For AGC, the CSM 220 isolates times during which local speech at a terminal 210 is thought to be present and then adjusts the signal gain so that the speech is near a  
5 predetermined level when considered on average. In one embodiment, the CSM 220 stores data in the room model 222 indicating the amount of signal gain to apply to the associated microphone 216.

For noise reduction, the CSM 220 estimates a frequency dependent profile of the noise floor and attenuates sounds which are temporarily near or below that noise floor. In  
10 one embodiment, the CSM 220 stores the estimated frequency dependent profile of a noise floor for a terminal in the associated room model 222.

Once the signal processing is performed 314, the CSM 220 uses the mixing module 226 to formulate 316 output mixes of the signal. If only one terminal 210 is producing an audio signal, and N is the number of terminals 210 engaged in the  
15 communications session, then the mixing module 226 formulates N-1 output mixes, where each audio mix is associated with a terminal other than the terminal from which the audio signal was received. In this case, the output mix merely contains the processed signal produced by the SP modules 224. If more than one terminal 210 is simultaneously producing an audio signal, then the mixing module 226 formulates N output mixes -- one  
20 mix for each terminal. In this latter case, each terminal's mix contains all of the received audio signals except for any audio signal received from the terminal for which the mix was formulated. Since a preferred embodiment of the present invention digitally mutes those terminals at which the CSM 220 determines no one is currently speaking, no

processing power is required to mix signals from those terminals. In addition, the output mixes can optionally be further processed according to the room model 222 for the associated terminal 210.

Next, the CSM 220 uses the SP modules 224 to encode 318 each mix into the appropriate format for its associated terminal 210. Since the encoding process can introduce noise or other artifacts into the signal, a preferred embodiment of the present invention decodes 320 each encoded mix and provides decoded mix samples to the associated room model 222. These samples become the sample values used to estimate the echo returning through the microphone 216 of that terminal when performing AEC.

Alternative embodiments of the present invention, however, use samples made before the mix is encoded. The CSM 220 sends 322 each mix to its associated terminal via the communications links 214. The terminal 210 plays the mix out of that terminal's speaker 218.

A preferred embodiment of the present invention supports easy upgrades of the processing switch 212 modules. The CSM 220, room models 222, and SP processing 224 modules are preferably software-upgradeable. In addition, one embodiment of the switch can be upgraded by replacing or adding modules. For example, the switch 212 can be upgraded by adding SP modules. In this manner, the performance of the switch 212 can be improved without altering the terminals 210 or otherwise incurring additional cost.

The above description is included to illustrate the operation of the preferred embodiments and is not meant to limit the scope of the invention. The scope of the invention is to be limited only by the following claims. From the above discussion, many

variations will be apparent to one skilled in the relevant art that would yet be encompassed by the spirit and scope of the invention.

002160" 50209960